

TITLE OF THE INVENTION

Video Encoding Method, Video Decoding Method, Video
Encoding Apparatus, Video Decoding Apparatus, Video
Encoding Program, and Video Decoding Program

5 BACKGROUND OF THE INVENTIONField of the Invention

[0001] The present invention relates to a video
encoding method, a video decoding method, a video
encoding apparatus, a video decoding apparatus, a video
10 processing system, a video encoding program, and a
video decoding program.

Related Background Art

[0002] Video signal encoding techniques are used
for transmission and storage-regeneration of video
15 signals. The well-known techniques include, for
example, the international standard video coding
methods such as ITU-T Recommendation H.263 (hereinafter
referred to as H.263), ISO/IEC International Standard
14496-2 (MPEG-4 Visual, hereinafter referred to as
20 MPEG-4), and so on. Another known newer encoding
method is a video coding method scheduled for joint
international standardization by ITU-T and ISO/IEC;
ITU-T Recommendation H.264 and ISO/IEC International
Standard 14496-10 (Joint Final Committee Draft of Joint
25 Video Specification, hereinafter referred to as H.26L).

[0003] Since a motion video signal consists of a

series of images (frames) varying little by little with time, it is common practice in these video coding methods to implement interframe prediction between a frame retrieved as a target for encoding (current
5 frame) and another frame (reference frame) and thereby reduce temporal redundancy in the video signal. In this case, where the interframe prediction is carried out between the current frame and a reference frame less different from the current frame, the redundancy
10 can be reduced more and encoding efficiency can be increased.

[0004] For this reason, as shown in Fig. 6, the reference frame for the current frame A1 can be either a temporally previous frame A0 or a temporally
15 subsequent frame A2 with respect to the current frame A1. The prediction with the previous frame is referred to as forward prediction, while the prediction with the subsequent frame as backward prediction. Bidirectional prediction is defined as a prediction in which one is
20 arbitrarily selected out of the two prediction methods, or as a prediction in which both methods are used simultaneously.

[0005] In general, with use of such bidirectional prediction, as in the example shown in Fig. 6, a
25 temporally previous frame as a reference frame for forward prediction and a temporally subsequent frame as

a reference frame for backward prediction each are preliminarily stored prior to the current frame.

[0006] Figs. 7A and 7B are diagrams showing (A) decoding and (B) output of the frames in the case of the bidirectional prediction shown in Fig. 6. For example, in the decoding of MPEG-4, where the current frame A1 is decoded by bidirectional interframe prediction, frame A0 being one temporally previous frame and frame A2 being one temporally subsequent frame with respect to the current frame A1 are first decoded as frames decoded by intraframe prediction without use of interframe prediction or as frames decoded by forward interframe prediction, prior to decoding of the current frame A1, and they are retained as reference frames. Thereafter, the current frame A1 is decoded by bidirectional prediction using these two frames A0, A2 thus retained (Fig. 7A).

[0007] In this case, therefore, the order of decoding times of the temporally subsequent reference frame A2 and the current frame A1 is reverse to the order of output times of their respective decoded images. Each of these frames A0, A1, and A2 is attached with output time information 0, 1, or 2, and thus the temporal sequence of the frames can be known according to this information. For this reason, the decoded images are outputted in the right order (Fig.

7B). In MPEG-4, the output time information is described as absolute values.

[0008] Some of the recent video coding methods permit the foregoing interframe prediction to be carried out using multiple reference frames, instead of one reference frame in the forward direction and one reference frame in the backward direction, so as to enable prediction from a frame with a smaller change from the current frame, as shown in Fig. 8. Fig. 8 shows an example using two temporally previous frames B0, B1 and two temporally subsequent frames B3, B4 with respect to the current frame B2, as reference frames for the current frame B2.

[0009] Figs. 9A and 9B are diagrams showing (A) decoding and (B) output of the frames in the case of the bidirectional prediction shown in Fig. 8. For example, in the decoding of H.26L, a plurality of reference frames can be retained within a range up to a predetermined upper bound of the number of reference frames and, on the occasion of carrying out interframe prediction, an optimal reference frame is arbitrarily designated out of them. In this case, where the current frame B2 is decoded as a bidirectionally predicted frame, the reference frames are first decoded prior to the decoding of the current frame B2; the reference frames include a plurality of temporally

previous frames (e.g., two frames B0, B1) and a plurality of temporally subsequent frames (e.g., two frames B3, B4) with respect to the current frame B2, which are decoded and retained as reference frames.

5 The current frame B2 can be predicted from a frame arbitrarily designated as the one used for prediction out of those frames B0, B1, B3, and B4 (Fig. 9A).

[0010] In this case, therefore, the order of decoding times of the temporally subsequent reference frames B3, B4 and the current frame B2 becomes reverse to the order of their respective output times. Each of these frames B0-B4 is attached with output time information or output order information 0-4, and the temporal sequence of the frames can be known according to this information. For this reason, the decoded images are outputted in the right order (Fig. 9B). The output time information is often described as absolute values. The output order is used where frame intervals are constant.

10

15

[0011] For carrying out the decoding by the backward prediction using temporally subsequent frames as predictive frames, it is necessary to satisfy the condition that the decoding of the temporally subsequent frames is completed prior to the decoding of the current frame so as to be available as predictive frames. In this case, a delay is incurred before the

20

25

decoded image of the current frame becomes available, as compared with a frame to which the backward prediction is not applied.

[0012] This will be specifically described below with reference to Figs. 10A to 10C. Figs. 10A to 10C correspond to the example shown in Figs. 6, 7A, and 7B. First, encoded data of each frame A0-A2 is decoded in an order necessary for execution of interframe prediction, and it is assumed that intervals of the frames are constant time intervals according to a frame rate and that the time necessary for the decoding operation is negligible for each frame A0-A2, regardless of whether the interframe prediction is applied and regardless of the directions of interframe prediction (Fig. 10A). In practice, the decoding intervals of the frames A0-A2 do not have to be constant and can change depending upon such factors as variation in encoding bits of the frames A0-A2 or the like; however, they can be assumed to be constant on average. The time necessary for the decoding operation is not zero, either, but it will raise no significant problem in the description hereinafter if the difference thereof is not so large among the frames A0-A2.

[0013] It is supposed herein that a time when a decoded image of frame A0 without delay due to backward

prediction and without reversal of the orders of decoding times and output times with respect to any other frame (a frame without delay and without reversal will be referred to hereinafter as a backward-prediction-nonassociated frame) is obtained, is defined
5 as an output time correlated with the decoded image, and the decoded image is outputted at the output time. Supposing the subsequent frame is the backward predicted frame A1, the decoded image thereof will be
10 decoded after the temporally subsequent frame A2, and a delay is thus made before the decoded image is obtained.

[0014] For this reason, if the time when the decoded image is obtained for the backward-prediction-nonassociated frame A0 is defined as a reference of
15 output time, the decoded image of the backward predicted frame A1 is not obtained by the output time correlated therewith (Fig. 10B). Namely, an output time interval between the decoded image of the
20 backward-prediction-nonassociated frame A0 and the decoded image of the backward predicted frame A1 becomes longer by the delay time necessary for execution of backward prediction than the original interval, which leads to unnatural video output.

25 [0015] Therefore, in the case where the backward interframe prediction is applied in video coding, as

shown in Fig. 10C, it is necessary to preliminarily delay the output time of the decoded image of the backward-prediction-nonassociated frame A0 by the delay time necessary for execution of the backward prediction as well so as to be able to correctly handle the output time interval to the backward predicted frame A1.

5 [0016] Conventionally, the backward interframe prediction was applied to video encoding under the conditions that encoding was carried out at a high bit rate and the fixed frame rate of 30 frames/second equal to that of TV broadcast signals was always used, like TV broadcasting or accumulation thereof, because backward interframe prediction brings about more options for prediction and hence increase of computational complexity so as to make implementation thereof difficult on simple equipment and because the increase of delay time was not desired in real-time communication involving bidirectional interlocution like video conferences.

20 [0017] In this case, for example, as in MPEG-4, where the use of one temporally subsequent frame as a reference frame for backward prediction, the delay time necessitated in execution of the backward prediction is constant. For example, where the frame rate is 30 frames/second as described above, the delay time is a time interval of each frame, i.e., $1/30$ second.

25

Accordingly, the time by which the output time of the decoded image of the backward-prediction-nonassociated frame should be delayed, can be equally set to 1/30 second.

5 SUMMARY OF THE INVENTION

[0018] In recent years, however, following the improvement in computer performance and progress in diversification of video services, delay is tolerable in video delivery through the Internet and mobile communications, and there is increased use of video coding requiring encoding at low bit rates. For implementing the encoding at low bit rates, frame rates smaller than 30 frames/second are applied, or variable frame rates are used to dynamically change the frame rate in order to control the encoding bit rate.

15 [0019] In such video coding, where the aforementioned backward prediction is applied in order to increase the encoding efficiency more, the delay time due to the backward prediction is not always 1/30 second as used before. In the application of variable frame rates, the frame rates are not constant. For example, in the case where a small frame rate is used on a temporary basis, the time interval of each frame there becomes large, and thus the time by which the output time of the decoded image of the backward-
25 prediction-nonassociated frame should be delayed is not

uniquely determined. For this reason, it becomes infeasible to correctly handle the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame.

5 [0020] In this case, there is such potential means that a large permissible delay time is preliminarily allowed for the backward prediction and that the output time of the decoded image of the backward-prediction-nonassociated frame is always delayed by this delay time, thereby correctly handling the output time interval relative to the decoded image of the backward predicted frame. In this case, however, the large delay is always added to the output time of the decoded image, regardless of the delay time in the practical backward prediction.

15 [0021] When multiple reference frames are used in the backward prediction as in H.26L, the decoding of all the reference frames being temporally subsequent frames must be completed prior to the decoding of the current frame. This further increases the delay time necessary for execution of the backward prediction.

20 [0022] In this case, since the number of reference frames used in the backward prediction is uniquely determined as a number of temporally subsequent frames to the current frame, which were decoded prior to the

25

current frame, the number of reference frames can be optionally changed within the range up to the predetermined upper bound of the maximum number of reference frames.

5 [0023] For example, supposing the upper bound of the number of reference frames is 4, the number of reference frames used in the backward prediction may be 2 as shown in Fig. 8, or 1 as shown in Fig. 11A, or 3 as shown in Fig. 11B. Since the number of reference frames can be changed in this way, the delay time necessary for execution of the backward prediction can vary largely. This leads to failure in correctly handling the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame.

10

15

[0024] At this time, since the maximum number of reference frames that can be used in the backward prediction does not exceed the upper bound of the number of reference frames, the delay time according to the upper bound of the number of reference frames is a maximum delay time that can be made in execution of the backward prediction. Therefore, if the output time of the decoded image of the backward-prediction-nonassociated frame is always delayed by this delay time, the output time interval relative to the decoded image of the backward predicted frame can be correctly

20

25

handled.

[0025] In this case, however, a large delay is always added to the output time of the decoded image, regardless of the number of reference frames actually used for the backward predicted frame. In the application of variable frame rates as described above, while the maximum number of reference frames can be uniquely determined, the maximum delay time cannot be uniquely determined.

10 [0026] In the application of the backward prediction to the video coding heretofore, it was infeasible to uniquely determine the delay time necessary for execution of the backward prediction, except for the case where use of a fixed frame rate was clear. This resulted in failure in correctly handling the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame, thus posing the problem that the video output became

20 unnatural.

[0027] In the case where multiple reference frames are used in the backward prediction, the number of reference frames can also be changed, so as to possibly vary the delay time. Therefore, there is the problem of the failure in correctly handling the time interval between the decoded image of the backward-prediction-

25

nonassociated frame and the decoded image of the backward predicted frame. In the case where the maximum delay time is always assumed in order to cope with this problem, there arises the problem that the large delay is always added to the output time of the decoded image.

[0028] The present invention has been accomplished in order to solve the above problems, and an object of the invention is to provide a video encoding method, a video decoding method, a video encoding apparatus, a video decoding apparatus, a video encoding program, and a video decoding program capable of achieving output of decoded images at appropriate time intervals when employing backward interframe prediction.

[0029] In order to achieve the above object, a video encoding method according to the present invention is a video encoding method of implementing interframe prediction between a frame and another frame, the video encoding method comprising: outputting a maximum delay time that is incurred by backward prediction.

[0030] Likewise, a video encoding apparatus according to the present invention is a video encoding apparatus for implementing interframe prediction between a frame and another frame, the video encoding apparatus being configured to: output a maximum delay

time that is incurred by backward prediction.

5 [0031] In the video encoding method and apparatus according to the present invention, as described above, on the occasion of encoding a moving picture consisting of a series of frames and outputting encoded data, the maximum delay time due to the backward prediction is outputted in addition to the encoded data. This enables achievement of output of decoded images at appropriate time intervals when employing the backward
10 interframe prediction.

[0032] A video encoding program according to the present invention is a video encoding program for letting a computer to execute video encoding of implementing interframe prediction between a frame and
15 another frame, the video encoding program letting the computer to execute: a process of outputting a maximum delay time that is incurred by backward prediction.

[0033] In the video encoding program according to the present invention, as described above, on the
20 occasion of encoding a moving picture and outputting encoded data thereof, the computer is made to execute the process of outputting the maximum delay time, in addition to the encoded data. This enables achievement of output of decoded images at appropriate time
25 intervals when employing the backward interframe prediction.

[0034] A video decoding method according to the present invention is a video decoding method of implementing interframe prediction between a frame and another frame, the video decoding method comprising:
5 effecting input of a maximum delay time that can be made by backward prediction.

[0035] Likewise, a video decoding apparatus according to the present invention is a video decoding apparatus for implementing interframe prediction
10 between a frame and another frame, the video decoding apparatus being configured to: effect input of a maximum delay time that is incurred by backward prediction.

[0036] In the video decoding method and apparatus
15 according to the present invention, as described above, on the occasion of decoding input encoded data to generate a moving picture, the maximum delay time due to the backward prediction is entered in addition to the encoded data. This enables achievement of output
20 of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0037] A video decoding program according to the present invention is a video decoding program for letting a computer to execute video decoding of
25 implementing interframe prediction between a frame and another frame, the video decoding program letting the

computer to execute: a process of effecting input of a maximum delay time that is incurred by backward prediction.

5 [0038] In the video decoding program according to the present invention, as described above, on the occasion of decoding encoded data to generate a moving picture, the computer is made to execute the process of effecting the input of the maximum delay time, in addition to the encoded data. This enables achievement of output of decoded images at appropriate time intervals when employing the backward interframe prediction.

15 [0039] Another video encoding method is one comprising an input step of effecting input of a frame as a target for encoding; an encoding step of encoding the frame by a predetermined method; and a maximum delay time calculating step of calculating a maximum delay time of the frame from a display time of the frame, an encoding time, and a delay time that is incurred by backward prediction.

20 [0040] Similarly, another video encoding apparatus is one comprising input means for effecting input of a frame as a target for encoding; encoding means for encoding the frame by a predetermined method; and maximum delay time calculating means for calculating a maximum delay time of the frame from a display time of

25

the frame, an encoding time, and a delay time that is incurred by backward prediction.

[0041] Similarly, another video encoding program is one for letting a computer to execute: an input process of effecting input of a frame as a target for encoding; an encoding process of encoding the frame by a predetermined method; and a maximum delay time calculating process of calculating a maximum delay time of the frame from a display time of the frame, an encoding time, and a delay time that is incurred by backward prediction.

[0042] In the video encoding method, apparatus, and program according to the present invention, as described above, the maximum delay time of the frame is calculated on the occasion of encoding a moving picture. This enables achievement of output of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0043] Another video decoding method is one comprising an input step of effecting input of image data containing encoded data of a frame encoded by a predetermined method, a decoding time of the frame, and a maximum delay time; a decoding step of decoding the encoded data to generate a regenerated image; and an image output time calculating step of calculating an output time for display of the frame, based on the

decoding time and the maximum delay time.

[0044] Similarly, another video decoding apparatus is one comprising input means for effecting input of image data containing encoded data of a frame encoded by a predetermined method, a decoding time of the frame, and a maximum delay time; decoding means for decoding the encoded data to generate a regenerated image; and image output time calculating means for calculating an output time for display of the frame, based on the decoding time and the maximum delay time.

[0045] Similarly, another video decoding program is one for letting a computer to execute: an input process of effecting input of image data containing encoded data of a frame encoded by a predetermined method, a decoding time of the frame, and a maximum delay time; a decoding process of decoding the encoded data to generate a regenerated image; and an image output time calculating process of calculating an output time for display of the frame, based on the decoding time and the maximum delay time.

[0046] In the video decoding method, apparatus, and program according to the present invention, as described above, on the occasion of decoding encoded data to generate a moving picture, the output time for display of the frame is calculated on the basis of the maximum delay time. This enables achievement of output

of decoded images at appropriate time intervals when employing the backward interframe prediction.

[0047] Concerning the maximum delay time outputted in the video encoding method, encoding apparatus, and encoding program, it is preferable to define the maximum delay time as a time difference between an occurrence time of a frame to be subjected to backward interframe prediction and an occurrence time of a temporally last subsequent frame that can be used as a reference frame in backward prediction.

[0048] Concerning application of the maximum delay time, the maximum delay time may be outputted as information to be applied to the entire encoded data. In another embodiment, the maximum delay time may be outputted as information to be applied to each frame. In still another embodiment, the maximum delay time may be optionally outputted as information to be applied to a frame for which the maximum delay time is indicated and to each temporally subsequent frame after the foregoing frame.

[0049] Concerning the maximum delay time entered in the video decoding method, decoding apparatus, and decoding program, it is preferable to define the maximum delay time as a time difference between a decoding time of a frame without reversal of orders of decoding times and output times with respect to any

other frame, and a decoded image output time correlated with the foregoing frame. In another embodiment, furthermore, it is preferable to set a reference for decoded image output times thereafter on the basis of the maximum delay time.

5 [0050] Concerning application of the maximum delay time, the maximum delay time may be entered as information to be applied to the entire encoded data. In another embodiment, the maximum delay time may be entered as information to be applied to each frame. In 10 still another embodiment, the maximum delay time may be optionally entered as information to be applied to a frame for which the maximum delay time is indicated and to each temporally subsequent frame after the foregoing frame.

15 [0051] A video processing system according to the present invention is a video processing system comprising a video encoding apparatus and a video decoding apparatus, wherein the encoding apparatus is the video encoding apparatus described above and 20 wherein the decoding apparatus is the video decoding apparatus described above.

[0052] As described above, the video processing system is constructed using the video encoding apparatus and the video decoding apparatus for 25 effecting output and input of the maximum delay time

due to the backward prediction. This substantializes the video processing system capable of achieving output of decoded images at appropriate time intervals when employing the backward interframe prediction.

5 [0053] The present invention will be more fully understood from the detailed description given hereinbelow and the accompanying drawings, which are given by way of illustration only and are not to be considered as limiting the present invention.

10 [0054] Further scope of applicability of the present invention will become apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating preferred
15 embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will be apparent to those skilled in the art from this detailed description.

20 BRIEF DESCRIPTION OF THE DRAWINGS

[0055] Fig. 1 is a block diagram showing the schematic structure of the video encoding apparatus, video decoding apparatus, and video processing system.

25 [0056] Fig. 2 is a diagram showing an example of encoding of frames in the case of the bidirectional prediction being carried out.

[0057] Fig. 3 is a block diagram showing an example of the configuration of the video encoding apparatus.

[0058] Fig. 4 is a block diagram showing an example of the configuration of the video decoding apparatus.

[0059] Figs. 5A and 5B are diagrams showing (A) decoding and (B) output of frames in the case of the bidirectional prediction shown in Fig. 2 being carried out.

[0060] Fig. 6 is a diagram showing encoding of frames in the case of the bidirectional prediction being carried out.

[0061] Figs. 7A and 7B are diagrams showing (A) decoding and (B) output of frames in the case of the bidirectional prediction shown in Fig. 6 being carried out.

[0062] Fig. 8 is a diagram showing encoding of frames in the case of the bidirectional prediction being carried out.

[0063] Figs. 9A and 9B are diagrams showing (A) decoding and (B) output of frames in the case of the bidirectional prediction shown in Fig. 8 being carried out.

[0064] Figs. 10A to 10C are diagrams showing (A) decoding, (B) output, and (C) delayed output of frames

in the case of the bidirectional prediction being carried out.

[0065] Figs. 11A and 11B are diagrams showing encoding of frames in the case of the bidirectional prediction being carried out.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0066] The preferred embodiments of the video encoding method, video decoding method, video encoding apparatus, video decoding apparatus, video encoding program, and video decoding program according to the present invention will be described below in detail with reference to the drawings. The same elements will be denoted by the same reference symbols throughout the description of the drawings, without redundant description thereof.

[0067] First, the encoding and decoding of moving picture in the present invention will be schematically described. Fig. 1 is a block diagram showing the schematic structure of the video encoding apparatus, video decoding apparatus, and video processing system according to the present invention. The video processing system is comprised of video encoding apparatus 1 and video decoding apparatus 2. The video encoding apparatus 1, video decoding apparatus 2, and video processing system will be described below together with the video encoding method and video

decoding method executed therein.

[0068] The video encoding apparatus 1 is a device configured to encode video data D0 consisting of a series of images (frames) and output encoded data D1, for transmission, for storage and regeneration of moving pictures. The video decoding apparatus 2 is a device configured to decode input encoded data D1 to generate decoded moving picture data D2 consisting of a series of frames. The video encoding apparatus 1 and the video decoding apparatus 2 are connected by a predetermined wired or wireless data transmission line, in order to transmit necessary data such as the encoded data D1 and others.

[0069] In the encoding of the moving picture carried out in the video encoding apparatus 1, as described previously, the interframe prediction is carried out between a frame of video data D0 entered as a target for encoding, and another frame as a reference frame, thereby reducing the redundancy in the video data. In the video processing system shown in Fig. 1, the video encoding apparatus 1 carries out the backward interframe prediction from a temporally subsequent frame for interframe prediction. Furthermore, this video encoding apparatus 1 outputs the maximum delay time that is incurred by the backward prediction, in addition to the encoded data D1.

[0070] In correspondence to such video encoding apparatus 1, the video decoding apparatus 2 is configured to effect input of the maximum delay time that is incurred by the backward prediction, in addition to the encoded data D1 from the video encoding apparatus 1. Then the video decoding apparatus 2 decodes the encoded data D1 with reference to the input maximum delay time to generate the video data D2.

[0071] By the video encoding apparatus 1 and video encoding method configured to output the maximum delay time, the video decoding apparatus 2 and video decoding method configured to effect input of the maximum delay time, and the video processing system equipped with those apparatus 1, 2, which are adapted for the backward interframe prediction as described above, it becomes feasible to achieve output of decoded images at appropriate time intervals in execution of the interframe prediction using the backward interframe prediction.

[0072] Concerning the maximum delay time outputted in the video coding, for example, the maximum delay time can be defined as a time difference between an occurrence time of a frame to be subjected to the backward interframe prediction and an occurrence time of a temporally last subsequent frame that can be used as a reference frame for backward prediction.

[0073] As for the maximum delay time entered in the video decoding, for example, the maximum delay time (hereinafter referred to as `dpb_output_delay`) can be defined as a time difference between a decoding time of a frame without delay due to backward interframe prediction and without reversal of orders of decoding times and output times with respect to other frame (the decoding time will be referred to hereinafter as T_r) and a decoded image output time correlated with the pertinent frame (the output time will be referred to hereinafter as T_o). In this case, preferably, a reference for decoded image output times thereafter is set based on the maximum delay time.

[0074] Application of the maximum delay time can be a method of applying it to entire encoded data or a method of applying it to each frame. Another application method is a method of applying the maximum delay time to each of the frames subsequent to the announcement of the information of the maximum delay time, i.e., to the frame for which the maximum delay time is indicated and to each of the frames temporally subsequent to that frame. The output, input, application, etc. of the maximum delay time in these methods will be specifically detailed later.

[0075] The processing corresponding to the video encoding method executed in the foregoing video

encoding apparatus 1 can be substantialized by the video encoding program for letting a computer to execute the video coding. The processing corresponding to the video decoding method executed in the video decoding apparatus 2 can be substantialized by the video decoding program for letting a computer to execute the video decoding.

[0076] For example, the video encoding apparatus 1 can be constructed of a CPU connected to a ROM storing software programs necessary for respective operations of the video coding and a RAM temporarily saving data during execution of a program. In this configuration, the video encoding apparatus 1 can be substantialized by letting the CPU to execute the predetermined video encoding program.

[0077] Similarly, the video decoding apparatus 2 can be constructed of a CPU connected to a ROM storing software programs necessary for respective operations of the video decoding and a RAM temporarily saving data during execution of a program. In this configuration, the video decoding apparatus 2 can be substantialized by letting the CPU to execute the predetermined video decoding program.

[0078] The above-stated program for letting the CPU to execute the processes for video encoding or for video decoding can be distributed in a form in which it

is recorded in a computer-readable recording medium. Such recording media include, for example, magnetic media such as hard disks and floppy disks, optical media such as CD-ROM and DVD-ROM, magnetooptical media such as floptical disks, or hardware devices, for example, such as RAM, ROM, and semiconductor nonvolatile memories, specially mounted to execute or store program commands.

5
10
15
20
[0079] The video encoding apparatus, the video decoding apparatus, the video processing system provided therewith shown in Fig. 1, and the video encoding method and video decoding method corresponding thereto will be described with specific embodiments. The description hereinafter will be based on the presumption that the encoding and decoding operations of motion video are implemented based on H.26L, and parts not specifically described about the operation in video encoding will be pursuant to the operation in H.26L. It is, however, noted that the present invention is not limited to H.26L.

[0080] (First Embodiment)

25
[0081] First, the first embodiment of the present invention will be described. The present embodiment will describe an embodied form of encoding at a fixed frame rate. In the encoding according to the present embodiment, the maximum number of reference frames used

for backward prediction is first determined, the maximum delay time is calculated thereafter from this maximum number of reference frames and the frame rate used in encoding, and the maximum delay time is then
5 outputted. In the decoding according to the present embodiment, on the occasion of decoding a backward-prediction-nonassociated frame, an output time of a decoded image thereof is delayed by the input maximum delay time. The delay time for the output time is
10 uniformly applied to every frame thereafter, so as to prevent the output time interval between the decoded image of the backward-prediction-nonassociated frame and the decoded image of the backward predicted frame from deviating from the original interval.

15 [0082] In the encoding, since the upper bound of the number of reference frames used is preliminarily determined, the maximum number of reference frames used for backward prediction is first determined within the range not exceeding the upper bound. Then, based on
20 the frame rate used for encoding, which is also preliminarily determined, the maximum delay time is calculated as a time interval of one frame or two or more frames according to the maximum number of reference frames used for backward prediction.

25 [0083] Fig. 2 is a diagram showing an example of encoding of a frame in execution of bidirectional

prediction. Here this Fig. 2 shows the example in which reference frames used for the current frame F2 are two temporally previous frames F0, F1 before the current frame F2 and two temporally subsequent frames F3, F4 after the current frame F2.

[0084] In the case where the maximum number of reference frames used for backward prediction is 2 and where the frame rate is 15 frames/second, as shown in Fig. 2, the time interval of one frame is $1/15$ second. In this case, therefore, the maximum delay time is $2 \times (1/15) = 2/15$ second.

[0085] In the encoding operation, encoding of each frame hereinafter is controlled so as not to carry out backward prediction requiring a delay time over the maximum delay time. Specifically, a sequence of encoding of frames is controlled so that any reference frame used in backward prediction, i.e., any temporally subsequent frame after the current frame is not encoded and outputted prior to the current frame over the maximum number of reference frames used in backward prediction.

[0086] Fig. 3 is a block diagram showing an example of the configuration of the video encoding apparatus used in the present embodiment. The video encoding apparatus 1 shown in Fig. 3 is comprised of an encoder 10 for encoding a frame (image) by the

predetermined method, a controller (CPU) 15 for
controlling operations of respective parts in the
encoding apparatus 1, a frame memory 11 disposed
between input terminal 1a and encoder 10, and a
5 multiplexer 12 disposed between output terminal 1b and
encoder 10. The controller 15 has a maximum delay time
calculator 16 for calculating the maximum delay time,
as a function thereof. The encoder 10 is provided with
an output buffer 13.

10 [0087] In the video encoding in the present
encoding apparatus 1, conditions for encoding of video
are entered through input terminal 1c. In this entry
of the conditions, the encoding conditions are
generally selected or entered through an input device
15 such as a keyboard. In the present embodiment,
specifically, the encoding conditions entered include
the size of a frame as a target for encoding, the frame
rate, and the bit rate and, in addition thereto, the
encoding conditions also include a predictive reference
20 structure of the video (whether backward prediction is
applied), the number of frames temporarily stored and
used as reference frames (corresponding to the capacity
of output buffer 13), and the number of reference
frames used in backward prediction. These conditions
25 may be set so as to vary with time. The encoding
conditions entered through the input terminal 1c are

stored into the controller 15.

[0088] With a start of the encoding operation, the controller 15 sends the encoding conditions to the encoder 10, where the encoding condition are set. On the other hand, a frame as an encoded object is entered through the input terminal 1a and is fed through the frame memory 11 to the encoder 10 to be encoded therein. The input frame is temporarily saved in the frame memory 11, because the order of frames is changed for execution of backward prediction. For example, in the example shown in Fig. 2, frame F2 is entered through the input terminal 1a before frames F3, F4, but it is encoded after the frames F3, F4; therefore, the frame F2 is temporarily saved in the frame memory 11.

[0089] The encoder 10 encodes the frame on the basis of the algorithm of H.26L. Then the encoded data is fed to the multiplexer 12 to be multiplexed with other related information and then the multiplexed data is outputted through the output terminal 1b. The frame used for the prediction is reproduced in the encoder 10 and is stored as a reference frame for encoding of the next frame into the buffer 13.

[0090] In the present embodiment, the maximum delay time calculator 16 of the controller 15 calculates the maximum delay time `dpb_output_delay`, based on the number of reference frames and the frame

rate entered through the input terminal 1c and used for the backward prediction. Then the multiplexer 12 adds the maximum delay time to encoded image data. In addition, an identifier (N) indicating a display turn for identification of each frame is also added together to the encoded data of each frame.

[0091] It is a matter of course that when the backward prediction is not applied, the number of reference frames used is zero and thus the value of dpb_output_delay is zero.

[0092] It is assumed in the present embodiment that a syntax for transmitting the maximum delay time is added to the encoded data syntax in H.26L, in order to implement the output of the maximum delay time in the encoding and the input of the maximum delay time in the decoding. In this example the new syntax is added into the Sequence Parameter Set being a syntax for transmitting the information to be applied to the entire encoded data.

[0093] The parameter dpb_output_delay is defined as a syntax for carrying the maximum delay time. It is assumed here that the parameter dpb_output_delay uses the same time unit used in the other syntaxes indicating the time in H.26L and that it indicates the maximum delay time in the time unit of 90 kHz. A numeral indicated in the time unit is encoded and

transmitted by a 32-bit unsigned fixed-length code.
For example, where the maximum delay time is $2/15$
second as described above, `dpb_output_delay` is $(2/15) \times$
 $90000 = 12000$.

5 [0094] In the decoding operation, the maximum
delay time carried by `dpb_output_delay` is decoded, and
an output time of a decoded image is delayed using it.

10 [0095] Fig. 4 is a block diagram showing an
example of the configuration of the video decoding
apparatus used in the present embodiment. The video
decoding apparatus 2 shown in Fig. 4 is comprised of a
decoder 20 for decoding encoded data to generate a
regenerated image, a controller (CPU) 25 for
controlling operations of respective parts in the
15 decoding apparatus 2, an input buffer 21 disposed
between input terminal 2a and decoder 20, and an output
buffer 22 disposed between output terminal 2b and
decoder 20. The controller 25 has an image output time
calculator 26 for calculating an output time for
20 display of a frame, as a function thereof.

25 [0096] In the video decoding in the present
decoding apparatus 2, data as a decoded object is
entered through the input terminal 2a. This data is
multiplexed data of the encoded data of each frame
encoded by the encoding apparatus 1 shown in Fig. 3,
the maximum delay time `dpb_output_delay`, and the

identifier (N) indicating the display turn of each frame.

[0097] The input data is stored into the input buffer 21. When a command from the controller 25 indicates arrival of a decoding time, data of one frame is entered from the input buffer 21 into the decoder 20 and is then decoded according to the algorithm of H.26L. The frame regenerated in this way is stored into the output buffer 22. The frame in the output buffer 22 is fed back via line 23 to decoder 20 to be used as a reference frame for decoding of the next frame.

[0098] On the other hand, the maximum delay time `dpb_output_delay`, the frame rate, and the identifier (N) of each frame decoded in the decoder 20 are fed into the controller 25. Then the image output time calculator 26 of the controller 25 calculates the output time of each frame from these data in accordance with the equation below.

$$To(n) = dpb_output_delay + N \times \text{frame interval}$$

In this equation, the frame interval is determined from the frame rate.

[0099] Supposing `dpb_output_delay` is 2/15 second and the frame interval is 1/15 second as in the example shown in Fig. 2, the output times of the respective frames are calculated as follows according to the above

equation.

$$N = 0, T_o(0) = 2/15$$

$$N = 1, T_o(1) = 3/15$$

$$N = 2, T_o(2) = 4/15$$

5 $N = 3, T_o(3) = 5/15$

According to the output times $T_o(n)$ obtained in this way by the controller 25, the frames in the output buffer 22 are outputted at constant intervals to the output terminal 2b, as indicated by frames F_0, F_1, F_2 , and F_3 shown in Fig. 5B. Although not illustrated, the output terminal 2b is connected to a display device such as a monitor.

[0100] Figs. 5A and 5B are diagrams showing (A) decoding and (B) output of the frames in the case of the bidirectional prediction shown in Fig. 2. It is assumed in the decoding operation that the encoded data of the frames is decoded in the order necessary for execution of the interframe prediction, the intervals thereof are constant time intervals according to the frame rate, and the time necessary for the decoding operation is negligible for each frame, regardless of whether interframe prediction is applied and regardless of the directions of interframe prediction. In this case, the maximum delay time necessary for execution of the backward prediction in the backward predicted frame is equal to a time interval of a frame or frames

according to the maximum number of reference frames used for the backward prediction. This time is carried as a maximum delay time by `dpb_output_delay`. Accordingly, for outputting a decoded image, an output time thereof is delayed by the maximum delay time.

[0101] In practice, the decoding intervals of the respective frames are not constant, and can vary depending upon such factors as variation in encoding bits of the frames. The time necessary for the decoding operation of each frame can also vary according to whether the frame is a backward predicted frame or according to encoding bits of each frame.

[0102] For delaying the output time, therefore, the reference is set at the time when the decoded image is obtained for the backward-prediction-nonassociated frame F0 without delay due to backward prediction and without reversal of orders of decoding times and output times with respect to any other frame, as shown in Figs. 5A and 5B. Namely, a time obtained by delaying the time when the decoded image is obtained, by the maximum delay time announced by `dpb_output_delay` is defined as a time equal to the output time correlated with this decoded image, and is used as a reference time in output of decoded images. The decoded images F1-F4 thereafter are outputted when this reference time agrees with a time equal to an output time correlated

with each decoded image.

[0103] For example, where the maximum delay time is 2/15 second as described above, a time at a delay of 2/15 second from the time when the decoded image is obtained for the backward-prediction-nonassociated frame, is defined as a time equal to the output time correlated with this decoded image and is used as a reference time in output of decoded images thereafter.

[0104] According to the circumstances, conceivably, the maximum delay time is not announced on purpose, in order to simplify the encoding or decoding operation. For such cases, the syntax for announcing the maximum delay time may be arranged to be omissible on the presumption that a flag to indicate the presence or absence of the syntax is transmitted prior to the syntax for transmitting the maximum delay time.

[0105] In the case where the announcement of the maximum delay time is omitted, the encoding operation may be preliminarily stipulated, for example, so as not to use the backward prediction or so that the number of reference frames used in backward prediction can be optionally altered within the range not exceeding the upper bound of the number of reference frames.

[0106] The decoding operation may be configured to perform in conformity with the stipulation in the encoding operation, for example, when backward

prediction is not applied, there occurs no delay
necessary for execution of backward prediction; or, the
decoding operation may also be configured so that the
number of reference frames used in backward prediction
5 can be optionally altered within the range not
exceeding the upper bound of the number of reference
frames, i.e., the delay time can vary large. In this
case, the decoding operation may be configured to
always perform processing assuming an expected maximum
10 delay time, or the decoding operation may be configured
to allow variation of output time intervals of decoded
images and perform simplified processing without
consideration to the delay time of each frame.

[0107] The present embodiment was described on the
15 assumption that the operations were implemented based
on H.26L, but it is noted that the video encoding
methods to which the present invention can be applied
are not limited to H.26L and that the present invention
can be applied to various video encoding methods using
20 the backward interframe prediction.

[0108] In the present embodiment, the syntax by
fixed-length codes was added as a syntax for
transmitting the maximum delay time into the Sequence
Parameter Set, but it is noted that the codes and
25 syntax for transmitting it, or the time unit for
expressing the maximum delay time are not limited to

these, of course. The fixed-length codes may be replaced by variable-length codes, and the maximum delay time can be transmitted by any of various syntaxes that can convey information to be applied to the entire encoded data.

5 [0109] For example, in H.26L, a syntax may be added into a Supplemental Enhancement Information Message. In a case using another video encoding method, the maximum delay time may be transmitted by a syntax for transmitting the information to be applied to the entire encoded data in the pertinent encoding method. In another case, the maximum delay time may also be transmitted outside the encoded data in the video encoding method as in ITU-T Recommendation H.245 used for conveying control information in communication using H.263.

15 [0110] (Second Embodiment)

[0111] The second embodiment of the present invention will be described below. The present embodiment will describe an embodied form of encoding at variable frame rates. The operations in the encoding and decoding according to the present embodiment are basically much the same as in the first embodiment. Since the present embodiment uses the variable frame rates, it involves an operation at low frame rates to avoid execution of the backward

prediction requiring the delay time over the
preliminarily calculated maximum delay time, in
addition to the operation in encoding in the first
embodiment, so as to prevent the output time interval
5 between the decoded image of the backward-prediction-
nonassociated frame and the decoded image of the
backward predicted frame from deviating from the
original interval even with variation of frame rates.

[0112] Since in the encoding operation the upper
10 bound of the number of reference frames is
preliminarily determined, the maximum number of
reference frames used for backward prediction is first
determined within the range not exceeding the upper
bound. Then the maximum frame time interval is
15 determined based on a target frame rate preliminarily
determined in control of encoding bit rates, and the
maximum delay time is calculated as a time interval of
one frame or two or more frames according to the
maximum number of reference frames used in backward
20 prediction and the maximum frame time interval.

[0113] In the encoding operation, encoding of each
frame thereafter is controlled so as to avoid the
backward prediction requiring the delay time beyond the
maximum delay time. Specifically, the order of
25 encoding of frames is controlled so as to prevent any
reference frame used in backward prediction, i.e., any

temporally subsequent frame after the current frame, that goes beyond the maximum number of reference frames used in backward prediction, from being encoded and outputted prior to the current frame.

5 [0114] In addition, when the encoding frame rate becomes temporarily small because of control of encoding bit rates, so as to make the frame time interval in that case larger than the maximum frame time interval, encoding of each frame is controlled so
10 as not to apply backward prediction to encoding of the frame there.

 [0115] The present embodiment is substantially identical to the first embodiment in that the maximum delay time is outputted in the encoding, in that the
15 syntax `dpb_output_delay` to transmit the maximum delay time is added to the encoded data syntax in order to effect input thereof in the decoding, and in the definition of the syntax.

 [0116] In the present embodiment, the decoding
20 operation is arranged to decode the maximum delay time announced by `dpb_output_delay` and delay the output time of the decoded image by use of it. This processing is also the same as in the first embodiment.

 [0117] (Third Embodiment)

25 [0118] The third embodiment of the present invention will be described below. The present

embodiment will describe an embodied form in which the maximum delay time is optionally announced for each frame and is thus flexibly changeable. The operations in the encoding and decoding according to the present
5 embodiment are basically similar to those in the first embodiment or the second embodiment.

[0119] In the present embodiment, the syntax `dpb_output_delay` to transmit the maximum delay time, which was defined in the first embodiment, is arranged
10 to be added into the Picture Parameter Set being a syntax to carry the information applied to each frame instead of the syntax to carry the information applied to the entire encoded data. The syntax `dpb_output_delay` herein is configured to indicate the
15 maximum delay time in the time unit of 90 kHz, as in the case of the first embodiment, and a numeral expressed in the time unit is encoded and transmitted by a 32-bit unsigned fixed-length code.

[0120] The present embodiment is much the same as
20 the first embodiment, as to the calculation of the maximum delay time in encoding and as to the delay of the output time of the decoded image by use of the maximum delay time in decoding. The configurations of the video encoding apparatus and video decoding
25 apparatus used in the present embodiment are much the same as those shown in Figs. 3 and 4 about the first

embodiment.

[0121] Let us explain how to determine the maximum delay time `dpb_output_delay` of each frame in the present embodiment. In the encoding apparatus 1 shown in Fig. 3, the controller 15 calculates the delay time (D) due to the backward prediction by the method as described in the first embodiment and determines the encoding time $Tr(n)$ of each frame. When a display time $Tin(n)$ of each frame is fed from frame memory 11, `dpb_output_delay(n)` of that frame is calculated as follows.

$$\text{dpb_output_delay}(n) = Tin(n) + D - Tr(n)$$

This value of `dpb_output_delay` is correlated with the pertinent frame and is multiplexed in the multiplexer 12.

[0122] In the present embodiment, the time $Tr(n)$ for encoding of each frame is also encoded together. Taking Fig. 2 as an example, $D = 2/15$ second, and $Tin(n) = 0, 1/15, 2/15, 3/15$, or $4/15$ ($n = 0, 1, 2, 3$, or 4). Because of change in the order of encoding, $Tr(n)$ becomes as follows: $Tr(n) = 0, 1/15, 4/15, 2/15$, or $3/15$ ($n = 0, 1, 2, 3$, or 4). Here `dpb_output_delay(n)` of each frame is obtained as follows.

$$\begin{aligned} n = 0, \text{ dpb_output_delay}(0) \\ = 0 + 2/15 - 0 = 2/15 \end{aligned}$$

$$\begin{aligned} n &= 1, \text{ dpb_output_delay}(1) \\ &= 1/15 + 2/15 - 1/15 = 2/15 \end{aligned}$$

$$\begin{aligned} n &= 2, \text{ dpb_output_delay}(2) \\ &= 2/15 + 2/15 - 4/15 = 0 \end{aligned}$$

$$\begin{aligned} 5 \quad n &= 3, \text{ dpb_output_delay}(3) \\ &= 3/15 + 2/15 - 2/15 = 3/15 \end{aligned}$$

$$\begin{aligned} n &= 4, \text{ dpb_output_delay}(4) \\ &= 4/15 + 2/15 - 3/15 = 3/15 \end{aligned}$$

[0123] On the other hand, in the decoding
 10 apparatus 2 shown in Fig. 4, the decoder 20 sends
 $\text{dpb_output_delay}(n)$ and $\text{Tr}(n)$ of each frame to the
 controller 25 and the controller 25 calculates the
 output time $\text{To}(n)$ of each frame on the basis of the
 equation below.

$$15 \quad \text{To}(n) = \text{Tr}(n) + \text{dpb_output_delay}$$

Taking Fig. 2 as an example, $\text{To}(n)$ for each frame is
 calculated as follows according to the above
 definition, based on $\text{Tr}(n) = 0, 1/15, 4/15, 2/15,$ or
 $3/15$ ($n = 0, 1, 2, 3,$ or 4) and $\text{dpb_output_delay}(n) =$
 20 $2/15, 2/15, 0, 3/15,$ or $3/15$ ($n = 0, 1, 2, 3,$ or 4).

$$n = 0, \text{ To}(0) = 0 + 2/15 = 2/15$$

$$n = 1, \text{ To}(1) = 1/15 + 2/15 = 3/15$$

$$n = 2, \text{ To}(2) = 4/15 + 0 = 4/15$$

$$n = 3, \text{ To}(3) = 2/15 + 3/15 = 5/15$$

$$25 \quad n = 4, \text{ To}(4) = 3/15 + 3/15 = 6/15$$

[0124] Namely, all the images are displayed with

the delay of 2/15 second and at constant intervals on the monitor. It is a matter of course that when the backward prediction is not applied, the number of reference frames used therefor is zero and the value of
5 dpb_output_delay(n) is thus zero.

[0125] Since the maximum delay time defines the reference time in output of decoded images from the time when the decoded image of the backward-prediction-nonassociated frame is acquired, it is enough to
10 transmit the maximum delay time only for the backward-prediction-nonassociated frame. It is therefore possible to employ, for example, a configuration wherein the syntax for transmitting the maximum delay time is arranged to be omissible on the presumption
15 that a flag indicating the presence or absence of the syntax is transmitted prior thereto. The syntax may also be arranged to be optionally omitted for the backward-prediction-nonassociated frame, provided that the maximum delay time transmitted before is applied in
20 that case where the maximum delay time is not transmitted.

[0126] The syntax for each frame in the present embodiment may also be used simultaneously with the syntax for the entire encoded data as defined in the
25 first embodiment. In this case, the syntax for each frame is omissible, provided that a flag indicating the

presence or absence of the syntax is transmitted prior thereto as described above. The maximum delay time transmitted in the syntax for the entire encoded data is continuously applied before the maximum delay time is transmitted in the syntax for each frame. After it is updated by the syntax for each frame, the time delayed based thereon is used as a reference time in output of every decoded image thereafter.

5 [0127] The present embodiment was described on the assumption that it was substantialized based on H.26L, but it is noted that the video encoding methods to which the present invention can be applied are not limited to H.26L and that the present invention can be applied to various video encoding methods using the backward interframe prediction.

15 [0128] In the present embodiment the syntax for transmitting the maximum delay time was the syntax by fixed-length codes added into the Picture Parameter Set, and it is a matter of course that the codes and syntax for transmitting it, or the time unit for expressing the maximum delay time are not limited to these, of course. The fixed-length codes can be replaced by variable-length codes, and the maximum delay time can be announced in any of various syntaxes capable of announcing the information to be applied to each frame.

20

25

[0129] For example, the syntax may be added into a Supplemental Enhancement Information Message in H.26L. When another video encoding method is applied, it is possible to use a syntax for announcing information to be applied to each frame in the pertinent encoding method. In addition, the information may also be announced outside the encoded data in the video encoding method as in ITU-T Recommendation H.245 used for announcement of control information in communication using H.263.

[0130] The video encoding method, video decoding method, video encoding apparatus, video decoding apparatus, video processing system, video encoding program, and video decoding program according to the present invention provide the following effect, as detailed above. Namely, when a moving picture consisting of a series of frames is encoded by the backward interframe prediction to be outputted, it becomes feasible to achieve output of decoded images at appropriate time intervals when employing the backward interframe prediction, by the video encoding method, encoding apparatus, and encoding program configured to output the maximum delay time due to the backward prediction, the video decoding method, decoding apparatus, and decoding program configured to effect input of the maximum delay time, and the video

processing system using them.

[0131] Particularly, different from the prior art, the output times are not absolute values, but relative values from the decoding time T_r ; therefore, the invention provides the effect of capability of accurately describing and transmitting the value of the maximum delay time `dpb_output_delay` by a small number of bits, even in the case that the frame rate is variable. Even if the decoding time T_r has a shift or is not received, a corresponding image will be outputted with a delay of `dpb_output_delay` from the time of completion of decoding, thus presenting the advantage that images are outputted at correct intervals.

[0132] From the invention thus described, it will be obvious that the invention may be varied in many ways. Such variations are not to be regarded as a departure from the spirit and scope of the invention, and all such modifications as would be obvious to one skilled in the art are intended for inclusion within the scope of the following claims.